

TITLE

METHOD FOR MUSIC ANALYSIS

BACKGROUND OF THE INVENTION

Field of the Invention:

5 The present invention relates to music analysis and particularly to a method for tempo estimation, beat detection and micro-change detection for music, which yields indices for alignment of soundtracks with video clips in an automated video editing system.

10 Description of the Related Art:

 Automatic extraction of rhythmic pulse from musical excerpts has been a topic of active research in recent years. Also called beat-tracking and foot-tapping, the goal is to construct a computational algorithm capable of
15 extracting a symbolic representation which corresponds to the phenomenal experience of "beat" or "pulse" in a human listener.

 "Rhythm" as a musical concept is intuitive to understand, but somewhat difficult to define. Handel writes "The
20 experience of rhythm involves movement, regularity, grouping, and yet accentuation and differentiation" Handel, 1989, p. 384 and also stresses the importance of the phenomenalist point of view- there is no ``ground truth`` for rhythm to be found in simple measurements of an acoustic
25 signal. The only ground truth is what human listeners agree to be the rhythmic aspects of the musical content of that signal.

As contrasted with "rhythm" in general, "beat" and "pulse" correspond only to "the sense of equally spaced temporal units" Handel, 1989 . Where "meter" and "rhythm" associate with qualities of grouping, hierarchy, and a strong/weak dichotomy, "pulses" in a piece of music are only periodic at a simple level. The beat of a piece of music is the sequence of equally spaced phenomenal impulses which define a tempo for the music.

It is important to note that there is no simple relationship between polyphonic complexity- the number and timbres of notes played at a single time- in a piece of music, and its rhythmic complexity or pulse complexity. There are pieces and styles of music which are texturally and timbrally complex, but have straightforward, perceptually simple rhythms; and there also exist musics which deal in less complex textures but are more difficult to rhythmically understand and describe.

The former sorts of musical pieces, as contrasted with the latter sorts, have a "strong beat". For these kinds of music, the rhythmic response of listeners is simple, immediate, and unambiguous, and every listener will agree on the rhythmic content.

In Automated Video Editing (AVE) systems, music analysis process is essential to acquire indices for alignment of soundtracks with video clips. In most pop music videos, video/image shot transitions usually occur at the beats. Moreover, fast music is usually aligned with many short video clips and fast transitions, while slow music is usually aligned with long video clips and slow transitions. Therefore, tempo estimation and beat detection

are two major and essential processes in an AVE system. In addition to beat and tempo, another important information essential to the AVE system is micro-changes, which is locally significant changes in a music, especially for music without drums or difficult to accurately detect beats and estimate tempo.

SUMMARY OF THE INVENTION

The object of the present invention is to provide a method for tempo estimation, beat detection and micro-change detection for music, which yields indices for alignment of soundtracks with video clips.

The present invention provides a method for music analysis comprising the steps of acquiring a music soundtrack, re-sampling an audio stream of the music soundtrack so that the re-sampled audio stream is composed of blocks, applying Fourier Transformation to each of the blocks, deriving a first vector from each of the transformed blocks, wherein components of the first vector are energy summations of the block within a plurality of first sub-bands, applying auto-correlation to each sequence composed of the components of the first vectors of all the blocks in the same first sub-band using a plurality of tempo values, wherein, for each sequence, a largest correlation result is identified as a confidence value and the tempo value generating the largest correlation result is identified as an estimated tempo, and comparing the confidence values of all the sequences to identify the estimated tempo corresponding to the largest confidence value as a final estimated tempo.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will become more fully understood from the detailed description given hereinbelow and the accompanying drawings, given by way of illustration only and thus not intended to be limitative of the present invention.

FIG. 1 is a flowchart of a method for tempo estimation, beat detection and micro-change detection according to one embodiment of the invention.

FIG. 2 shows the audio blocks according to one embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

FIG.1 is a flowchart of a method for tempo estimation, beat detection and micro-change detection according to one embodiment of the invention.

In step S10, a music soundtrack is acquired. For example, the tempo of the music soundtrack ranges from 60 to 180 M.M.(beats per minute).

In step S11, the audio stream of the music soundtrack is preprocessed. The audio stream is re-sampled. As shown in FIG. 2, the original audio stream is divided into chunks C1, C2,..., each including, for example, 256 samples. The block B1 is composed of the chunks C1 and C2, the block B2 is composed of the chunks C2 and C3, and so forth. Thus, the blocks B1, B2,... have samples overlapping with each other.

In step S12, FFT is applied to each audio block, which converts the audio blocks from time domain to frequency domain.

In step S13, a pair of sub-band vectors are derived from each audio block, wherein one vector is for tempo estimation and beat detection while the other is for micro-change detection. The components of each vector are energy summations of the audio block within different frequency ranges (sub-bands) and the sub-band sets for the two vectors are different. The vectors may be represented by:

$$V1_{(n)} = (A_1(n), A_2(n), \dots, A_I(n)) \text{ and} \\
V2_{(n)} = (B_1(n), B_2(n), \dots, B_J(n)) ,$$

where $V1_{(n)}$ and $V2_{(n)}$ are the two vectors derived from the n^{th} audio block, $A_i(n)$ ($i=1\sim I$) is the energy summation of the n^{th} audio block within the i^{th} sub-band of the sub-band set for tempo estimation and beat detection, and $B_j(n)$ ($j=1\sim J$) is the energy summation of the n^{th} audio block within the j^{th} sub-band of the sub-band set for micro-change detection. Further, the energy summations are derived from the following equations:

$$A_i(n) = \sqrt{\sum_{k=L_i}^{H_i} a(n,k)} \text{ and} \\
B_j(n) = \sqrt{\sum_{k=L_j}^{H_j} a(n,k)} ,$$

where L_i and H_i are the lower and upper bounds of the i^{th} sub-band of the sub-band set for tempo estimation and beat detection, L_j and H_j are the lower and upper bounds of the j^{th} sub-band of the sub-band set for micro-change detection, and $a(n,k)$ is the energy value (amplitude) of the n^{th} audio block at frequency k . For example, the sub-band set for tempo estimation and beat detection comprises three sub-bands [0Hz, 125Hz], [125Hz, 250Hz] and [250Hz, 500Hz] while

that for micro-change detection comprises four sub-bands
[0Hz, 1100Hz], [1100Hz, 2500Hz], [2500Hz, 5500Hz] and
[5500Hz, 11000Hz]. Since drum sounds with low frequencies
are so regular in most pop music that beat onsets can be
5 easily derived from them, the total range of the sub-band
set for tempo estimation and beat detection is lower than
that for micro-change detection.

In step S141, each sequence composed of the components
in the same sub-band of the vectors $V1_{(1)}$, $V1_{(2)}$, ..., $V1_{(N)}$ (N
10 is the number of the audio blocks) is filtered to eliminate
noise. For example, there are three sequences respectively
for the sub-bands [0Hz, 125Hz], [125Hz, 250Hz] and [250Hz,
500Hz]. In each sequence, only the components having
amplitudes larger than a predetermined value are left
15 unchanged while the others are set to zero.

In step S142, auto-correlation is applied to each of
the filtered sequences. In each filtered sequence,
correlation results are calculated using tempo values, for
example, from 60 to 186 M.M., wherein the tempo value
20 generating the largest correlation results is the estimated
tempo and a confidence value of the estimated tempo is the
largest correlation results. Additionally, a threshold for
determination of validity of the correlation results may be
used, wherein only the correlation results larger than the
25 threshold is valid. If there is no valid correlation
results in one of the sub-bands, the estimated tempo and
confidence value of that sub-band are set to 60 and 0
respectively.

In step S143, by comparing the confidence values of the
30 estimated tempo of all the sub-bands for tempo estimation

and beat detection, the estimated tempo with the largest confidence value is determined as the final estimated tempo.

5 In step S144, the beat onsets are determined by the final estimated tempo. First, the maximum peak in the sequence of the sub-band whose estimated tempo is the final estimated tempo is identified. Second, the neighbors of the maximum peak within a range of the final estimated tempo is deleted. Third, the next maximum peak in the sequence is identified. Fourth, the second and third steps are repeated
10 until no more peak is identified. These identified peaks are beat onsets.

In step 15, micro-changes in the music soundtrack is detected using the sub-band vectors $V2_{(1)}$, $V2_{(2)}$, ..., $V2_{(N)}$. A micro-change value MV is calculated for each audio block.
15 The micro-change value is the sum of differences between the current vector and previous vectors. More specifically, the micro-change value of the nth audio block is derived by the following equation:

$$MV_{(n)} = \text{Sum}(\text{Diff}(V2_{(n)}, V2_{(n-1)}), \text{Diff}(V2_{(n)}, V2_{(n-2)}), \text{Diff}(V2_{(n)}, V2_{(n-3)}), \text{Diff}(V2_{(n)}, V2_{(n-4)}))$$

20 The difference between two vectors may be defined variously. For example, it may be the difference between the amplitudes of the two vectors. After the micro-change values are derived, they are compared to a predetermined threshold. The audio blocks having micro-change values larger than the
25 threshold are identified as micro-changes.

In the previously described embodiment, the sub-band sets may be determined by user input, which achieves an interactive music analysis.

In conclusion, the present invention provides a method
30 for tempo estimation, beat detection and micro-change

detection for music, which yields indices for alignment of soundtracks with video clips. The tempo value, beat onsets and micro-changes are detected using sub-band vectors of audio blocks having overlapping samples. The sub-band sets
5 defining the vectors may be determined by user input. Thus, the indices for alignment of soundtracks with video clips are more accurate and easily derived.

The foregoing description of the preferred embodiments of this invention has been presented for purposes of
10 illustration and description. Obvious modifications or variations are possible in light of the above teaching. The embodiments were chosen and described to provide the best illustration of the principles of this invention and its practical application to thereby enable those skilled in the
15 art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. All such modifications and variations are within the scope of the present invention as determined by the appended claims when interpreted in accordance with the
20 breadth to which they are fairly, legally, and equitably entitled.